

## SAMPLING AND STATISTICAL METHODS FOR BEHAVIORAL ECOLOGISTS

This book describes the sampling and statistical methods used most often by behavioral ecologists and field biologists. Written by a biologist and two statisticians, it provides a rigorous discussion, together with worked examples, of statistical concepts and methods that are generally not covered in introductory courses, and which are consequently poorly understood and applied by field biologists. The first section reviews important issues such as defining the statistical population when using non-random methods for sample selection, bias, interpretation of statistical tests, confidence intervals and multiple comparisons. After a detailed discussion of sampling methods and multiple regression, subsequent chapters discuss specialized problems such as pseudoreplication, and their solutions. It will quickly become *the* statistical handbook for all field biologists.

JONATHAN BART is a Research Wildlife Biologist, in the Biological Resources Division of the U.S. Geological Survey. He is currently based at the Snake River Field Station of the Forest and Rangeland Ecosystem Science Center in Boise, Idaho.

MICHAEL A. FLIGNER is Professor and Vice Chair of the Department of Statistics at the Ohio State University.

WILLIAM I. NOTZ is also a Professor in the Department of Statistics at the Ohio State University, and has served as Associate Editor for the *Journal of the American Statistical Association*, *Technometrics* and the *Journal of Statistical Education*.

To Susan  
Alysha and Karen  
Claudia

SAMPLING AND  
STATISTICAL METHODS FOR  
BEHAVIORAL ECOLOGISTS

JONATHAN BART, MICHAEL A. FLIGNER,  
AND WILLIAM I. NOTZ



PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE  
The Pitt Building, Trumpington Street, Cambridge CB2 1RP, United Kingdom

CAMBRIDGE UNIVERSITY PRESS  
The Edinburgh Building, Cambridge CB2 2RU, UK <http://www.cup.cam.ac.uk>  
40 West 20th Street, New York, NY 10011-4211, USA <http://www.cup.org>  
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

© Cambridge University Press 1998

This book is in copyright. Subject to statutory exception  
and to the provisions of relevant collective licensing agreements,  
no reproduction of any part may take place without  
the written permission of Cambridge University Press.

First published 1998

Printed in the United Kingdom at the University Press, Cambridge

Typeset in Times NR 10/13 [SE]

*A catalog record for this book is available from the British Library*

*Library of Congress Cataloging in Publication data*

Bart, Jonathan.  
Sampling and Statistical Methods for Behavioral Ecologists / Jonathan Bart, Michael A.  
Fligner, William I. Notz.  
p. cm.  
Includes index.  
ISBN 0 521 45095 0 hardback.  
ISBN 0 521 45705 X paperback.  
1. Animal behavior – Statistical methods. 2. Sampling (Statistics)  
I. Fligner, Michael A. II. Notz, William I. III. Title.  
QL751.65.S73B37 1998  
590'.1'5195 dc21 97–35086 CIP

ISBN 0 521 45095 0 hardback  
ISBN 0 521 45705 X paperback

# Contents

	<i>Preface</i>	<i>page</i> ix
<b>1</b>	<b>Statistical analysis in behavioral ecology</b>	<b>1</b>
1.1	Introduction	1
1.2	Specifying the population	1
1.3	Inferences about the population	5
1.4	Extrapolation to other populations	11
1.5	Summary	12
<b>2</b>	<b>Estimation</b>	<b>14</b>
2.1	Introduction	14
2.2	Notation and definitions	15
2.3	Distributions of discrete random variables	17
2.4	Expected value	21
2.5	Variance and covariance	24
2.6	Standard deviation and standard error	26
2.7	Estimated standard errors	26
2.8	Estimating variability in a population	30
2.9	More on expected value	32
2.10	Linear transformations	34
2.11	The Taylor series approximation	36
2.12	Maximum likelihood estimation	42
2.13	Summary	45
<b>3</b>	<b>Tests and confidence intervals</b>	<b>47</b>
3.1	Introduction	47
3.2	Statistical tests	47
3.3	Confidence intervals	58
3.4	Sample size requirements and power	65
3.5	Parametric tests for one and two samples	68

3.6	Nonparametric tests for one or two samples	78
3.7	Tests for more than two samples	81
3.8	Summary	84
<b>4</b>	<b>Survey sampling methods</b>	<b>85</b>
4.1	Introduction	85
4.2	Overview	86
4.3	The finite population correction	97
4.4	Sample selection methods	99
4.5	Multistage sampling	109
4.6	Stratified sampling	124
4.7	Comparison of the methods	131
4.8	Additional methods	132
4.9	Notation for complex designs	137
4.10	Nonrandom sampling in complex designs	139
4.11	Summary	146
<b>5</b>	<b>Regression</b>	<b>148</b>
5.1	Introduction	148
5.2	Scatterplots and correlation	148
5.3	Simple linear regression	154
5.4	Multiple regression	159
5.5	Regression with multistage sampling	174
5.6	Summary	176
<b>6</b>	<b>Pseudoreplication</b>	<b>177</b>
6.1	Introduction	177
6.2	Power <i>versus</i> generality	178
6.3	Fish, fish tanks, and fish trials	182
6.4	The great playback debate	185
6.5	Causal inferences with unreplicated treatments	187
6.6	Summary	187
<b>7</b>	<b>Sampling behavior</b>	<b>190</b>
7.1	Introduction	190
7.2	Defining behaviors and bouts	190
7.3	Allocation of effort	192
7.4	Obtaining the data	196
7.5	Analysis	197
7.6	Summary	199

<b>8</b>	<b>Monitoring abundance</b>	<b>200</b>
8.1	Introduction	200
8.2	Defining ‘the trend’	201
8.3	Estimating standard errors	209
8.4	Outliers and missing data	210
8.5	Index methods	211
8.6	Pseudoreplication	216
8.7	Summary	217
<b>9</b>	<b>Capture–recapture methods</b>	<b>219</b>
9.1	Introduction	219
9.2	Rationale	219
9.3	Capture histories and models	220
9.4	Model selection	222
9.5	Closed population models	222
9.6	Open population models	223
9.7	Summary	226
<b>10</b>	<b>Estimating survivorship</b>	<b>228</b>
10.1	Introduction	228
10.2	Telemetry studies	228
10.3	Nesting success	231
10.4	Summary	236
<b>11</b>	<b>Resource selection</b>	<b>238</b>
11.1	Introduction	238
11.2	Population units and parameters	239
11.3	Several animals	243
11.4	Multivariate definition of resources	244
11.5	Summary	246
<b>12</b>	<b>Other statistical methods</b>	<b>248</b>
12.1	Introduction	248
12.2	Adaptive sampling	248
12.3	Line transect sampling	249
12.4	Path analysis	250
12.5	Sequential analysis	251
12.6	Community analysis	253
12.7	Summary	254

<b>APPENDIX ONE</b>	Frequently used statistical methods	257
<b>APPENDIX TWO</b>	Statistical tables	279
<b>APPENDIX THREE</b>	Notes for Appendix One	311
<i>References</i>		320
<i>Index</i>		328

## Preface

This book describes the sampling and statistical methods used most often by behavioral ecologists. We define behavioral ecology broadly to include behavior, ecology and such related disciplines as fisheries, wildlife, and environmental physiology. Most researchers in these areas have studied basic statistical methods, but frequently have trouble solving their design or analysis problems despite having taken these courses. The general reason for these problems is probably that introductory statistics courses are intended for workers in many fields, and each field presents a special, and to some extent unique, set of problems. A course tailored for behavioral ecologists would necessarily contain much material of little interest to students in other fields.

The statistical problems that seem to cause behavioral ecologists the most difficulty can be divided into several categories.

1. Some of the most difficult problems faced by behavioral ecologists attempting to design a study or analyze the resulting data fall between statistics – as it is usually taught – and biology. Examples include how to define the sampled and target populations, the nature and purpose of statistical analysis when samples are collected nonrandomly, and how to avoid pseudoreplication.
2. Some methods used frequently by behavioral ecologists are not covered in most introductory texts. Examples include survey sampling, capture–recapture, and distance sampling.
3. Certain concepts in statistics seem to need reinforcement even though they are well covered in many texts. Examples include the rationale of statistical tests, the meaning of confidence intervals, and the interpretation of regression coefficients.
4. Behavioral ecologists encounter special statistical problems in certain areas including index methods, detecting habitat ‘preferences’, and sampling behavior.

5. A few mathematical methods of use to behavioral ecologists are generally not covered in introductory methods courses. Examples include the statistical properties of ratios and other nonlinear combinations of random variables, rules of expectation, the principle of maximum likelihood estimation, and the Taylor series approximation.

This book is an attempt to address problems such as those above adopting the special perspective of behavioral ecology. Throughout the book, our general goals have been that behavioral ecologists would find the material relevant and that statisticians would find the treatment rigorous. We assume that readers will have taken one or more introductory statistics courses, and we view our book as a supplement, rather than a substitute, for these courses.

The book is based in part on our own research and consulting during the past 20 years. Before writing the text, however, we undertook a survey of the methods used by behavioral ecologists. We did this by examining every article published during 1990 in the journals *Behavioral Ecology and Sociobiology*, *Animal Behavior*, *Ecology*, and *The Journal of Wildlife Management* and all the articles on behavior or ecology published in *Science* and *Nature*. We tabulated the methods in these articles and used the results frequently in deciding what to include in the book and how to present the examples.

Chapter One describes statistical objectives of behavioral ecologists emphasizing how the statistical and nonstatistical aspects of data analysis reinforce each other. Chapter Two describes estimation techniques, introducing several statistical methods that are useful to behavioral ecologists. It is more mathematical than the rest of the book and can be skimmed by readers less interested in such methods. Chapter Three discusses tests and confidence intervals concentrating on the rationale of each method. Methods for ratios are discussed as are sample size and power calculations. The validity of *t*-tests when underlying data are non-normal is discussed in detail, as are the strengths and weaknesses of nonparametric tests. Chapter Four discusses survey sampling methods in considerable detail. Different sampling approaches are described graphically. Sample selection methods are then discussed followed by a description of multistage sampling and stratification. Problems caused by non-random sample selection are examined in detail. Chapter Five discusses regression methods emphasizing conceptual issues and how to use computer software to carry out general linear models' analysis.

The first five Chapters cover material included in the first few courses in statistical methods. In these Chapters, we concentrate on topics that

behavioral ecologists often have difficulty with, assuming that the reader has already been exposed to the basic methods and ideas. The subsequent Chapters discuss topics that are generally not covered in introductory statistics courses. We introduce each topic and provide suggestions for additional reading. Chapter Six discusses the difficult problem of pseudo-replication, introducing an approach which we believe might help to resolve the controversies in this area and focus the discussions on biological, rather than statistical, issues. Chapter Seven discusses special statistical problems that arise in sampling behavior. Chapter Eight discusses estimating and monitoring abundance, particularly by index methods. Chapter Nine discusses capture-recapture methods, while Chapter Ten emphasizes the estimation of survival. Chapter Eleven discusses resource selection and Chapter Twelve briefly mentions some other topics of interest to behavioral ecologists with suggestions for additional reading.

Appendix One gives a detailed explanation of frequently used statistical methods, whilst Appendix Two contains a set of tables for reference. They are included primarily so that readers can examine the formulas in more detail to understand how analyses are conducted. We have relegated this material to an appendix because most analyses are carried out using statistical packages and many readers will not be interested in the details of the analysis. Nonetheless, we encourage readers to study the material in the appendices as doing so will greatly increase one's understanding of the analyses. In addition, some methods (e.g., analysis of stratified samples) are not available in many statistical packages but can easily be carried out by readers able to write simple computer programs. Appendix Three contains detailed notes on derivation of the material in Appendix One.

This book is intended primarily for researchers who wish to use sampling techniques and statistical analysis as a tool but who do not have a deep interest in the underlying mathematical principles. We suspect, however, that many biologists will be interested in learning more about the statistical principles and techniques used to develop the methods we present. Knowledge of this material is of great practical use because problems arise frequently which can be solved readily by use of these methods, but which are intractable without them. Basic principles of expectation (by which many variance formulas may be derived) and use of the Taylor series approximation (by which nearly all the remaining variance formulas needed by behavioral ecologists may be derived) are examples of these methods. Maximum likelihood estimation is another statistical method that can be presented without recourse to complex math and is frequently of value to biologists. We introduce these methods in Chapter Two and

illustrate their use periodically in the rest of the book. These sections, however, can be skipped without compromising the reader's ability to understand later sections of the book.

Another approach of great utility in developing a deep understanding of the statistical methods we present is to prepare computer programs that carry out calculations and simulations. We encourage readers to learn some programming in an elementary language such as Basic or the languages included in many data bases or statistical packages and then to write short programs to investigate the material we present. Several opportunities for such projects are identified in the text, and all of the examples we mention are listed in the *Index* under the heading 'Computer programming, examples'. We have found that preparing programs in this manner not only ensures that one understands the fine structure of the analysis, but in addition frequently leads one to think much more deeply about how the statistical analysis helps us understand natural systems. Such efforts also increase one's intuition about whether studies can be carried out successfully given the resources available and about how to allocate resources among different segments of the study. Furthermore, data management, while not discussed in this book, frequently consumes far more time during analysis than carrying out the actual statistical tests, and in many studies is nearly impossible without recourse to computer programs. For all of these reasons, we encourage readers strongly to learn a programming language.

The authors thank the staff of Cambridge University Press for their assistance with manuscript preparation, especially our copy editor, Sarah Price. Much of the book was written while the senior author was a member of the Zoology Department at Ohio State University. He acknowledges the many stimulating discussions of biological statistics with colleagues there, especially Susan Earnst, Tom Grubb, and John Harder and their graduate students. JB also acknowledges his intellectual debt to Douglas S. Robson of Cornell University who introduced him to sampling techniques and other branches of statistics and from whom he first learned the value of integrating statistics and biology in the process of biological research.

# 1

## Statistical analysis in behavioral ecology

### 1.1 Introduction

This Chapter provides an overview of how statistical problems are formulated in behavioral ecology. We begin by identifying some of the difficulties that behavioral ecologists face in deciding what population to study. This decision is usually made largely on nonstatistical grounds but a few statistical considerations are worth discussing. We then introduce the subject of making inferences about the population, describing objectives in statistical terms and discussing accuracy and the general ways used to measure it. Finally, we note that statistical inferences do not necessarily apply beyond the population sampled and emphasize the value of drawing a sharp distinction between the sampled population and larger populations of interest.

### 1.2 Specifying the population

Several conflicting goals influence decisions about how large and variable the study population should be. The issues are largely nonstatistical and thus outside the scope of this book, but a brief summary, emphasizing statistical issues insofar as they do occur, may be helpful.

One issue of fundamental importance is whether the population of interest is well defined. Populations are often well defined in wildlife monitoring studies. The agencies carrying out such studies are usually concerned with a specific area such as a State and clearly wish to survey as much of the area as possible. In observational studies, we would often like to collect the data throughout the daylight hours – or some portion of them – and throughout the season we are studying.

Sampling throughout the population of interest, however, may be difficult for practical reasons. For example, restricting surveys to roads and

observations to one period of the day may permit the collection of a larger sample size. A choice then arises between ‘internal and external validity’. If surveys are restricted to roadsides, then smaller standard errors may be obtained, thereby increasing ‘internal validity’, but we will worry that trends along the roads may differ from trends for the entire area, thus reducing ‘external validity’. A similar problem may occur if observations are restricted to certain times of day or portions of the season. When the population of interest is well defined, as in these cases, then the trade-off between internal and external validity is conceptually straightforward, though deciding how to resolve it in specific cases may be difficult.

When there is no single well-defined population of interest, then the situation is a little more complex conceptually. Consider the following example. Suppose we are investigating the relationship between dominance and time spent watching for predators in groups of foraging animals. Dominant individuals might spend more time foraging because they assume positions of relative security from predators. Alternatively, they might spend less time foraging because they obtain better foraging positions and satisfy their nutritional requirements more quickly. Suppose that we can study six foraging groups in one woodlot, or two groups in each of three woodlots. Sampling three woodlots might seem preferable because the sampled population would then be larger and presumably more representative of the population in the general area. But suppose that dominant individuals spend more time foraging in some habitats and less time foraging in others. With three woodlots – and perhaps three habitats – we might not obtain statistically significant differences between the foraging time of dominants and subdominants due to the variation among woodlots. We might also not have enough data within woodlots to obtain statistically significant effects. Thus, we would either reach no conclusion or, by averaging over woodlots, incorrectly conclude that dominance does not affect vigilance time. This unfortunate outcome might be much less likely if we confined sampling to a single woodlot. Future study might then show that the initial result was habitat dependent.

In this example, there is no well-defined target population about which we would like to make inferences. The goal is to understand an interesting process. Deciding how general the process is can be viewed as a different goal, to be undertaken in different studies. Thus, while the same trade-off between internal and external validity occurs, there is much less of a premium on high external validity. If the process occurs in the same way across a large population, and if effort can be distributed across this population without too much reduction in sample sizes, due to logistic

costs, then having a relatively large sampled population may be worthwhile. But if such a plan increases logistic costs, or if the process varies across the population, then restricting the population in space, time or other ways may be preferable.

Studies conducted in one location or within 1 year are sometimes criticized on the grounds that the sample size is 1. In some sense, however, nearly all studies have a sample size of 1 because they are carried out in one county, state, or continent. Frequently, those arguing for a distribution of the study across two or more areas or years are really arguing that two or more complete studies should have been conducted. They want enough data to determine whether the results hold in each area or year. This is desirable of course. Two studies are nearly always better than one; but, if the sample size is only sufficient to obtain one good estimate, then little may be gained, and much lost, by spreading the effort over a large area or long period of time.

### *Superpopulations*

Sometimes a data set is collected without any formal random selection – this occurs in many fields. In behavioral ecology, it is most likely when the study is conducted within a well-defined area and all individuals (typically plants or animals) within the boundaries of the area are measured. It might be argued that in such cases we have taken a census (i.e., measured all members) of the population so that calculation of standard errors and statistical tests is neither needed nor appropriate. This view is correct if our interest really is restricted to individuals in the study area at the time of the study. In the great majority of applications, however, we are really interested in an underlying process, or at least a much larger population than the individuals we studied.

In sampling theory, a possible point of view is that many factors not under our control operate in essentially a random manner to determine what individuals will be present when we do our study, and that the individuals present can thus be regarded as a random sample of the individuals that might have been present. Such factors might include weather conditions, predation levels, which migrants happened to land in the area, and so on. In sampling theory, such hypothetical populations are often called ‘superpopulations’ (e.g., Cochran 1977 p. 158; Kotz and Johnson 1988). We assume that our sample is representative of the superpopulation and thus that statistical inferences apply to this larger group of individuals. If the average measurement from males, for example, is significantly larger than the average from females, then we may legitimately conclude that the average for all males that

might have been present in our study area probably exceeds the average for all females. If the difference is not significant, then the data do not support any firm conclusion about which sex has the larger average value. Note that asserting the existence of a superpopulation, describing the individuals it contains, and modeling its relation to our sample require biological or ecological arguments as much as or more than statistical arguments.

The superpopulation concept can also be explained by reference to an ‘assignment process’. The word assignment refers to the underlying biological process, not to randomization carried out by the investigator. To illustrate the concept, imagine that we are comparing survival rates of males and females. We might view the individuals of each sex as being ‘assigned’ to one of two groups at the end of the study, alive and dead, and the process may be viewed as having random elements such as whether a predator happens to encounter a given individual. The question is whether members of one sex are more likely than the other to be assigned to the ‘alive’ group. The superpopulation is then the set of possible outcomes and inferences apply to the underlying probabilities of survival for males and females. This rationale is appealing because it emphasizes our interest in the underlying process, rather than in the individuals who happened to be present when we conducted the study.

Justifying statistical analysis by invoking the superpopulation concept might be criticized on the basis that there is little point in making inferences about a population if we cannot clearly describe what individuals comprise the population. There are two responses to this criticism. First, there is an important difference between deciding whether sample results might have arisen by chance and deciding how widely conclusions from a study apply. In the example above, if the sample results are not significantly different then we have not shown that survival rates are sex specific for *any* population (other than the sample we measured). The analysis thus prevents our making unwarranted claims. Second, describing the sampled population, in a particular study, is often not of great value even if it is possible. The main value of describing the sampled population is that we can then generalize the results from our sample to this population. But in biological research, we usually want to extend our findings to other areas, times, and species, and clearly the applicability of our results to these populations can only be determined by repeating the study elsewhere. Thus, the generality of research findings is established mainly by repeating the study, not by precisely demarcating the sampled population in the initial study.

Statisticians tend to view superpopulations as an abstraction, as opposed to a well-defined population about which inferences are to be made.

Behavioral ecologists thus must use care when invoking this concept to ensure that the rationale is reasonable. For example, one would probably not measure population size in a series of years and then declare that the years could be viewed as a random sample from a superpopulation of years. Population size at one time often depends strongly on population size in recent years so consecutive years could not legitimately be viewed as an independent sample. Nonetheless, in many studies in the field of behavioral ecology we can imagine much larger populations which we suspect our samples are representative of and to which we would like to make inferences. In such cases statistical analysis is appropriate because it helps guard against unwarranted conclusions.

### 1.3 Inferences about the population

#### *Objectives*

Although biologists study a vast array of species, areas, behaviors, and so on, most of the parameters estimated may be assigned to a small number of categories. Most quantities of interest in behavioral ecology are of two types: (1) means, proportions, or quantities derived from them, such as differences; and (2) measures of association such as correlation and regression coefficients and the quantities based on them such as regression equations. Estimates of these quantities are often called 'point estimates'. In addition, we usually want an estimate of accuracy such as a standard error. A point estimate coupled with an estimate of accuracy can often be used to construct a confidence interval or 'interval estimate', an interval within which we are relatively confident the true parameter value lies. Frequent use is made later in the book of the phrase 'point and interval estimates'.

#### *Definitions*

One of the first steps in obtaining point or interval estimates is to clearly understand the statistical terms. In behavioral ecology, the connection between the terms and the real problem is sometimes surprisingly difficult to specify, as will become clear later in the book. Here we introduce a few terms and provide several examples of how they would be defined in different studies.

The quantity we are trying to estimate is referred to as a parameter. Formally, a parameter is any numerical characteristic of a population. In estimating density, the parameter is actual density in the sampled population. In estimating change in density, the parameter is change in the actual

densities. The term random variable refers to any quantity whose numerical value depends on which sample we happen to obtain by random selection. The sample mean is thus a random variable as is any quantity calculated from the sample such as a standard deviation or standard error.

A numerical constant is typically a *known* quantity that is not of direct interest and whose value does not depend on the particular sample selected. For example, if we estimate density per m<sup>2</sup> but then multiply the estimate by 10,000 to obtain density per hectare, then the 10,000 is a numerical constant. On the other hand, a parameter is an *unknown* constant whose value does not depend on the particular sample selected but is of direct interest.

In any analysis, one must identify the units in the sample and the measurements taken on each unit. Thus, we may define the sample mean, with respect to some variable as  $\bar{y} = \sum y_i / n$  where  $n$  is the sample size and  $y_i$ ,  $i = 1, \dots, n$  are the measurements. In this book, we generally follow the tradition of survey sampling in which a distinction is made between the population units and the variables measured on each unit in the sample. Population units are the things we select during random sampling; variables are the measurements we record.

If we capture animals and record their sex, age, and mass, then the population unit is an animal and the variables are sex, age, and mass. If we record behavioral measurements on each of several animals during several 1-h intervals, then the population unit is an animal watched for 1 h, an 'animal-hour', and the variables are the behavioral data recorded during each hour of observation. In time-activity sampling, we often record behavior periodically during an observation interval. The population unit is then an 'animal-time', and the variables are the behaviors recorded. In some studies, plants or animals are the variables rather than the population units. For example, if we record the number of plants or the number of species in each of several plots, then the population unit is a plot, and the variable is 'number of plants' or 'number of species'. In most studies carried out by behavioral ecologists, the population unit is: (1) an animal, plant, or other object; (2) a location in space such as a plot, transect, or dimensionless point; (3) a period or instant of time; or (4) a combination involving time such as an animal watched for 1 h or a location sampled at each of several times.

Nearly all sampling plans assume that the population units are nonoverlapping. Usually this can be accomplished easily in behavioral ecology. For example, if the population units are plots, then the method of selecting the plots should ensure that no two plots in the sample will overlap each other. In some sampling plans, the investigator begins by dividing the population

units into groups in such a way that each population unit is in one and only one group. Subdivision in this manner is called a partition of the population. Sample selection is also usually assumed to be without replacement unless stated otherwise. Sampling without replacement implies that a unit cannot be selected twice for the sample, while units could be included two or more times when sampling is with replacement. The names are derived from the practice of physically removing objects from the population, as in drawing balls from an urn and then replacing them or not replacing them.

Application of the ‘population unit/variable’ approach may seem difficult at first in estimating proportions. If we select ‘ $n$ ’ plants and record the proportion that have flowers, what is ‘the variable’? Statisticians usually approach such problems by defining the population unit as an individual and the variable as 0 if the individual does not have the trait or condition of interest and 1 if it does. The proportion is thus the mean of the variables in the sample. For example, let  $y_i$  refer to the  $i^{\text{th}}$  plant ( $i = 1, \dots, n$ ) and equal 0 if the plant does not have flowers and 1 if it does have flowers. Then the proportion may be written as  $\sum y_i / n$ . This principle – that proportions may be thought of as means (of 0s and 1s) – is useful in several contexts. For example, it shows that all results applicable to means in general also apply to proportions (though proportions do have certain properties – described in later Chapters – not shared by all means). Notice that it matters whether we use 0 to mean ‘a plant without flowers’ or ‘a plant with flowers’. The term ‘success’ is commonly used to indicate which category is identified by a 1. The other category is often called ‘failure’. In our example, a ‘success’ would mean a plant with flowers.

In most studies we wish to estimate many different quantities, and the definitions of population units and variables may change as we calculate new estimates. For example, suppose we wish to estimate the average number of plants/m<sup>2</sup> and seeds/plant. We use plots to collect plants and then count the number of seeds on each plant. In estimating the average number of plants per plot, the population unit is a plot, and the variable is the number of plants (i.e.,  $y_i$  = the number of plants in the  $i^{\text{th}}$  plot). In estimating the number of seeds per plant, the population unit is a plant, and the variable is the number of seeds (i.e.,  $y_i$  = the number of seeds on the  $i^{\text{th}}$  plant).

The population is the set of all population units that might be selected for inclusion in the sample. The population has the same ‘dimensions’ as the population units. If a population unit is an animal watched for an observation interval, then, by implication, the population has two dimensions, one for the animals that might be selected, the other for the times that might be

selected. The population in this case might be envisaged as an array, with animals that might be selected listed down the side and times that might be selected listed across the top. Cells in the array thus represent population units and the entries in them are the variables. This approach of visualizing the population as a two-dimensional array will be used extensively in our discussions of ‘Survey sampling methods’ (Chapter Four) and ‘Pseudo-replication’ (Chapter Six).

Biologists often think of the species as ‘the population’ they are studying. The statistical population, however, is the set of population units that might enter the sample. If the population units are plots (in which we count animals for instance), then the statistical population is a set of plots. If the population unit is a trap left open for a day, then the statistical population is the set of trap-days that might enter the sample, not the animals that we might catch in them. This is just a matter of semantics, but confusion is sometimes avoided by distinguishing between statistical and biological populations.

#### *Measures of error*

The term error, in statistics, has approximately the same meaning as it does in other contexts: an estimate likely to be far from the true value has large error and one likely to be close to the true value has small error. Two kinds of error, sampling error and bias, are usually distinguished. The familiar ‘bull’s eye’ analogy is helpful to explain the difference between them. Imagine having a quantity of interest (the bull’s eye) and a series of estimates (individual bullets lodged on the target). The size of the shot pattern indicates sampling error and the difference, if any, between the center of the shot pattern and the bull’s eye indicates bias. Thus, sampling error refers to the variation from one sample to another; bias refers to the difference (possibly zero) between the mean of all possible estimates and the parameter.

Notice that the terms sampling error and bias refer to the pattern that would be observed in repeated sampling, not to a single estimate. We use the term estimator for the method of selecting a sample and analyzing the resulting data. Sampling error and bias are said to be properties of the estimator (e.g., we may say the estimator is biased or unbiased). Technically, it is not correct to refer to the bias or sampling error of a single estimate. More important than the semantics, however, is the principle that measures of error reveal properties of the set of all possible estimates. They do not automatically inform us about how close the single estimate we obtain in a real study is to the true value. Such inferences can be made but the reasoning is quite subtle. This point, which must be grasped to understand the

rationale of statistical inference, is discussed more in Chapter Three, 'Tests and confidence intervals'.

The quantity most widely used to describe the magnitude of sampling error is called the standard error of the estimate. One of the remarkable properties of modern statistical methods is that standard errors – a measure of the variation that would occur in repeated sampling – can usually be estimated from a single sample. The effects of sampling error can also be described by the coefficient of variation ( $CV$ ) which expresses the standard error as a percentage of the estimate [i.e.,  $CV = (\text{standard error}/\text{estimate}) \times 100\%$ ]. Calculation of  $CV$  values facilitates comparison of estimates, especially of quantities measured on very different scales. For example, an investigator might report that all the  $CV$  values were less than 20%. Sampling error is also sometimes measured by the variance of the estimate, which is the square of the standard error.

Three sources of bias may be distinguished: selection bias, measurement bias, and statistical bias. Selection bias may occur when some units in the population are more likely to be selected than others or are selected but not measured (but the investigator is using a procedure which assumes equally likely selection probabilities). Measurement bias is the result of systematic recording errors. For example, if we are attempting to count all individuals in plots but usually miss some of those present, then our counts are subject to measurement bias. Note that measurement errors do not automatically cause bias. If positive and negative errors tend to balance, then the average value of the error in repeated sampling might be zero, in which case no measurement bias is present. Statistical bias arises as a result of the procedures used to analyze the data and the statistical assumptions that are made.

Most statistical textbooks do not discuss selection and measurement bias in much detail. In behavioral ecology, however, it is often unwise to ignore these kinds of error. Selection of animals for study must often be done using nonrandom sampling, so selection bias may be present. In estimating abundance, we often must use methods which we know do not detect every animal. Many behavioral or morphological measurements are difficult to record accurately, especially under field conditions.

The statistical bias of most commonly used statistical procedures is either zero or negligible, a condition we refer to as 'essentially unbiased', meaning that the bias, while not exactly equal to zero, is not of practical importance. When using newer statistical procedures, especially ones developed by the investigator, careful study should be given to whether statistical bias exists. When estimates are biased, then upper bounds must be placed

on the size of the bias or the estimates are of little value. This is often possible using analytical methods for statistical bias. Bias caused by nonrandom selection or measurement errors, however, usually cannot be estimated with statistical methods, a point which has important implications for understanding tests and confidence intervals (see Chapter Three).

A few examples will help clarify the distinctions between sampling error and the various types of bias. Leuschner *et al.* (1989) selected a simple random sample of hunters in the southeastern United States of America and asked them whether more tax dollars should be spent on wildlife. The purpose was to estimate what proportion of all hunters in the study area would answer yes to this question. Sampling error was present in the study because different random samples of hunters would contain different proportions who felt that tax dollars should be spent on wildlife. Selection bias could have been present because 42% of the people selected for the sample were unreachable, gave unusable answers, or did not answer at all. These people might have felt differently, as a group, than those who did answer the question. There is no reason to believe that measurement bias was present. The authors used standard, widely accepted methods to analyze their results, so it is unlikely that their estimation procedure contained any serious statistical bias. Note that the types of error are distinct from one another. Stating, as in the example above, that no measurement or statistical bias was present in the estimates does not reveal anything about the magnitude of sampling error or selection bias.

Otis *et al.* (1978) developed statistical procedures for estimating population size when animals are captured, marked, and released, and then some of them are recaptured one or more times. The quantity of interest was the total number of animals in the population (assumed in these particular models to remain constant during the study). Sampling error would occur because the estimates depend on which animals are captured and this in turn depends on numerous factors not under the biologists' control. Selection bias could occur if certain types of animals were more likely to be captured than others (though the models allowed for certain kinds of variation in capture probabilities). In the extreme case that some animals are so 'trap wary' as to be uncapturable, these animals would never appear in any sample. Thus, the estimator would estimate the population size of capturable animals only and thus systematically underestimate total population size. Measurement bias would occur if animals lost their marks (this was assumed not to occur). The statistical procedures were new, so the authors studied statistical bias with computer simulations. They found

little statistical bias under some conditions, but under other conditions the estimates were consistently too high or too low even if all required assumptions were met.

Two other terms commonly used to describe the different components of error are precision and accuracy. Precision refers solely to sampling error whereas accuracy refers to the effects of both sampling error and bias. Thus, an estimator may be described as ‘precise but not accurate’ meaning it has a small standard error but is biased. Accuracy is defined as the square of the standard error plus the square of the bias and is also known as the mean squared error of the estimator.

#### 1.4 Extrapolation to other populations

Statistical analysis allows us to make rigorous inferences about the statistical population but does not automatically allow us to make inferences to any other or larger population. By ‘statistical population’ we mean the population units that might have entered the sample. When measurements are complex or subjective, then the scope of the statistical inferences may also be limited to the ‘conditions of the study’, meaning any aspect of the study that might have affected the outcome. These restrictions are often easy to forget or ignore in behavioral ecology so here we provide a few examples.

If we record measurements from a series of animals in a study area, then the sampled population consists of the animals in the study area at the time of the study and the statistical inferences apply to this set of animals. If we carry out a manipulation involving treatments and controls, then ‘the population’ is the set of individuals that might have been selected and the inferences apply only to this population and experiment. Inferences about results that would have been obtained with other populations or using other procedures may be reasonable but they are not justified by the statistical analysis. With methods that detect an unknown fraction of the individuals present (i.e., index methods), inferences apply to the set of outcomes that might have been obtained, not necessarily to the biological populations, because detection rates may vary. Attempts to identify causes in observational studies must nearly always recognize that the statistical analysis identifies differences but not the cause of the differences.

One sometimes hears that extrapolation beyond the sampled population is ‘invalid’. We believe that this statement is too strong, and prefer saying that extrapolation of conclusions beyond the sampled population must be

based on additional evidence, and that this evidence is often largely or entirely nonstatistical. This does not mean that conclusions about a target population are wrong: it only means that the protection against errors afforded by the initial statistical methods is not available and everyone should realize that. For example, if we measure clutch size in one study area and period of time, then the statistical analysis only justifies making inferences about the birds in the study area during the study period. Yet everyone would agree that the results tell us a good deal about likely clutch size in nearby areas and in future or past years. The extent to which conclusions from the study can be extrapolated to larger target populations would be evaluated using biological information such as how clutch size varies in space and time in the study species and other closely related species. This distinction is often reflected in the organization of journal articles. The Results section contains the statistical analysis, whereas analyses of how widely the results apply elsewhere are presented in the Discussion section. Thus, in our view, the reason for careful identification of the sampled population and conditions of the study is not to castigate those who extrapolate conclusions of the study beyond this population but only to emphasize that additional, and usually nonstatistical, rationales must be developed for this stage of the analysis.

### **1.5 Summary**

Decisions about what population to study are usually based primarily on practical, rather than statistical, grounds but it may be helpful to recognize the trade-off between internal and external validity and to recognize that studying a small population well is often preferable to studying the largest population possible. The superpopulation concept helps explain the role of statistical analysis when all individuals in the study area have been measured. Point estimates of interest in behavioral ecology usually are means or measures of association, or quantities based on them such as differences and regression equations. The first step in calculating point estimates is defining the population unit and variable. A two-dimensional array representing the population is often helpful in defining the population. Two measures of error are normally distinguished: sampling error and bias. Both terms are defined with respect to the set of all possible samples that might be obtained from the population. Sampling error is a measure of how different the sample outcomes would be from each other. Bias is the difference between the average of all possible outcomes and the quantity of interest, referred to as the parameter. Three types of bias may

be distinguished: selection bias, measurement bias, and statistical bias. Most statistical methods assume the first two types are absent but this is often not a safe assumption in behavioral ecology. Statistical inferences provide a rigorous method for drawing conclusions about the sampled population, but inferences to larger populations must be based on additional evidence. It is therefore useful to distinguish clearly between the sampled population and larger target populations of interest.